Proceedings of the 10th World Congress of EWRA 'Panta Rhei', 5-9 July 2017, Athens, Greece

Forecasting of geophysical processes using stochastic and machine learning algorithms

G.A. Papacharalampous*, H. Tyralis and D. Koutsoyiannis

[1] *Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Polytechniou 5, 157 80 Zografou, Greece*

* *e-mail: papacharalampous.georgia@gmail.com*

Abstract: We perform an extensive comparison between four stochastic and two machine learning (ML) forecasting algorithms by conducting a multiple-case study. The latter is composed by 50 single-case studies, which use time series of total monthly precipitation and mean monthly temperature observed in Greece. We apply a fixed methodology to each individual case and, subsequently, we perform a cross-case synthesis to facilitate the detection of systematic patterns. The stochastic algorithms include the Autoregressive order one model, an algorithm from the family of Autoregressive Fractionally Integrated Moving Average models, an Exponential Smoothing State Space algorithm and the Theta algorithm, while the ML algorithms are Neural Networks and Support Vector Machines. We also use the last observation as a Naive benchmark in the comparisons. We apply the forecasting methods to the deseasonalized time series. We compare the one-step ahead as also the multi-step ahead forecasting properties of the algorithms. Regarding the one-step ahead forecasting properties, the assessment is based on the absolute error of the forecast of the last observation. For the comparison of the multi-step ahead forecasting properties we use five metrics applied to the test set (last twelve observations), i.e. the root mean square error, the Nash-Sutcliffe efficiency, the ratio of standard deviations, the index of agreement and the coefficient of correlation. Concerning the ML algorithms, we also perform a sensitivity analysis for time lag selection. Additionally, we compare more sophisticated ML methods as regards to the hyperparameter optimization to simple ones.

Key words: Stochastic algorithms, machine learning algorithms, multiple-case study, cross-case synthesis, multi-step ahead forecasting, one-step ahead forecasting, time lag selection, hyperparameter optimization

1. INTRODUCTION

Machine learning (ML) algorithms are widely used for the forecasting of geophysical processes as an alternative to stochastic algorithms. Popular ML algorithms are the rather well established Neural Networks (NN) and the new-entrant in most scientific fields Support Vector Machines (SVM). The latter was presented in its current form by Cortes and Vapnik (1995; see also Vapnik, 1995, 1999). The large number of the relevant applications is imprinted in Maier and Dandy (2000) and Raghavendra and Deka (2014).

As a result, the research in geophysical sciences often focuses on comparing stochastic to ML forecasting algorithms. The comparisons performed are usually based on single-case studies (e.g. Koutsoyiannis et al., 2008; Valipour et al., 2013), which offer the benefit of studying the phenomena in detail as also in their context and thus can provide interesting insights. On the other hand, single-case studies do not allow generalizations in any extent (Achen and Snidal, 1989). Generalizations could be derived by examining a sufficient number of different cases, as implemented in Papacharalampous et al. (2017). Within the latter study large-scale computational experiments based on simulations are conducted to compare several stochastic and ML methods regarding their multi-step ahead forecasting properties. A statistical analysis is performed and the results are presented accordingly.

Here we conduct a multiple-case study composed by 50 individual cases, each of them based on geophysical time series data from Greece. We apply a fixed methodology to each individual case for the comparison between several stochastic and ML methods regarding their one-step ahead and multi-step ahead forecasting properties. Concerning the ML methods, we also perform a sensitivity analysis for time lag selection. Additionally, we compare more sophisticated ML methods as regards to the hyperparameter optimization to simple ones. Finally, we perform a cross-case synthesis to facilitate the detection of systematic patterns. We believe that the multiple-case study method can be useful for the comparative assessment of forecasting methods, as it can provide a form of generalization named "contingent empirical generalization", while retaining the immediacy of the single-case study method (Achen and Snidal, 1989).

2. DATA AND METHODS

2.1 Time series

We use 50 time series of total monthly precipitation and mean monthly temperature observed in Greece. We select only those with few missing values (blocks with length equal or less than one). Subsequently, we use the Kalman filter algorithm from the zoo R package (Zeileis and Grothendieck, 2005) for filling in the missing values. We use the deseasonalized time series for the application of the forecasting methods for the improvement of the forecasting quality, as suggested in Taieb et al. (2012). The deseasonalization is performed using a multiplicative model of time series decomposition.

The basic information about the time series is provided in Table 1. To describe the long-term persistence of the deseasonalized time series, we estimate the Hurst parameter H for each of them using the maximum likelihood method (Tyralis and Koutsoyiannis, 2011) implemented with the HKprocess R package (Tyralis, 2016).

2.2 Forecasting methods

We use four stochastic and two ML forecasting algorithms. The stochastic algorithms include the Autoregressive order one model (AR(1)), an algorithm from the family of Autoregressive Fractionally Integrated Moving Average models (auto_ARFIMA), an Exponential Smoothing State Space algorithm (BATS) and the Theta algorithm. The ML algorithms are Neural Networks (NN) and Support Vector Machines (SVM). We also use the last observation as a Naive benchmark in the comparisons. We apply the stochastic algorithms using the forecast R package (Hyndman and Khandakar, 2008; Hyndman et al., 2016) and the ML using the rminer R package (Cortez, 2010, 2015). The Naive, AR(1), auto_ARFIMA and BATS algorithms apply Box-Cox transformation to the input data before fitting a model to them.

While the stochastic forecasting methods are simply defined by the stochastic algorithm, the ML methods are defined by the set {ML algorithm, hyperparameter selection procedure, time lags}. We compare two procedures for hyperparameter selection, i.e. predefined hyperparameters or defined after optimization, and 21 regression matrices, each using the first n time lags, $n = 1, 2, \ldots, 21$. The hyperparameter optimization is performed with the hold-out method.

Hereafter, we consider that the ML models are used with predefined hyperparameters and that the regression matrix is built only by the first time lag, unless mentioned differently. We use two ML forecasting methods (one for each algorithm) in the comparisons conducted between stochastic and machine learning. We also use 42 forecasting methods (21 for each algorithm) to perform a sensitivity analysis for time lag selection and four ML forecasting methods (two for each algorithm) for the investigation of the effect of the hyperparameter optimization.

2.3 Metrics

Regarding the one-step ahead forecasting properties, the assessment is based on the absolute error (AE) of the forecast of the last observation. For the comparison of the multi-step ahead forecasting properties we use the Root Mean Square Error (RMSE), the Nash-Sutcliffe efficiency (NSE), the ratio of standard deviations (rSD), the index of agreement (d) and the coefficient of correlation (Pr) applied to the test set. These metrics quantify the forecasting methods' performance according to several criteria related to the accuracy, the capture of the variance and the correlation between the forecasted and their respective observed values. For the definitions of the metrics NSE, d and Pr the reader is referred to Krause et al. (2005), while for the definition of the rSD to Zambrano-Bigiarini (2014).

Table 1. Time series examined. The Hurst parameter H is estimated for the deseasonalized time series.

s/n	Process	Code	Location	Station id	Reference	Start	End	Length (months)	H
1	Precipitation	prec_1	Agrinion	16672000	Peterson and Vose (1997)	Jan 1956	Dec 1987	384	0.48
2		prec_2	Alexandroupoli	16627000		Jan 1951	Dec 1990	480	0.59
3		prec_3	Aliartos	16674000		Jan 1907	Dec 1990	1008	0.53
4		prec_4	Anogeia	16754001		Jan 1919	Dec 1939	252	0.52
5		prec_5	Anogeia	16754001		Jan 1950	Dec 1979	360	0.53
6		prec_6	Araxos	16687000		Jan 1949	Dec 2000	624	0.51
7		prec_7	Athens	16714000		Jan 1860	Dec 1881	264	0.48
8		prec_8	Athens	16714000		Jan 1887	Dec 2005	1428	0.53
9		prec_9	Athens	16716000		Jan 1929	Dec 1945	204	0.52
10		prec_10	Fragma	16715001		Jan 1926	Dec 1990	780	0.54
11		prec_11	Heraklion	16754000		Jan 1946	Dec 1990	540	0.50
12		prec_12	Igoumenitsa	16641001		Jan 1951	Dec 1990	480	0.49
13		prec_13	Ioannina	16642000		Jan 1951	Dec 1990	480	0.58
14		prec_14	Kalamata	16726000		Jan 1956	Dec 1970	180	0.51
15		prec_15	Kalo Chorio	16756001		Jan 1950	Dec 1984	420	0.50
16		prec_16	Kastelli	16760001		Jan 1949	Dec 1976	336	0.55
17		prec_17	Kerkyra	16641000		Jan 1952	Dec 1996	540	0.51
18		prec_18	Kythira	16743000		Jan 1951	Dec 1973	276	0.48
19		prec_19	Kos	16742000		Jan 1958	Dec 1990	396	0.49
20		prec_20	Kozani	16632000		Jan 1955	Dec 1987	396	0.57
21		prec_21	Larissa	16648000		Jan 1951	Dec 1997	564	0.55
22		prec_22	Lemnos	16650001		Jan 1951	Dec 2000	600	0.52
23		prec_23	Methoni	16734000		Jan 1951	Dec 1991	492	0.49
24		prec_24	Milos	16738000		Jan 1951	Dec 1990	480	0.57
25		prec_25	Mytilene	16667000		Jan 1952	Dec 1990	468	0.55
26		prec_26	Naxos	16732000		Jan 1955	Dec 1971	204	0.46
27		prec_27	Patra	16689000		Jan 1901	Dec 1984	1008	0.52
28		prec_28	Sitia	16757000		Jan 1960	Dec 1983	288	0.56
29		prec_29	Skyros	16684000		Jan 1955	Dec 1987	396	0.50
30		prec_30	Thessaloniki	16622000		Jan 1931	Dec 1997	804	0.58
31		prec_31	Thessaloniki	16622002		Jan 1961	Dec 1970	120	0.56
32		prec_32	Trikala	16645001		Jan 1951	Dec 1990	480	0.56
33		prec_33	Tripoli	16710000		Jan 1951	Dec 1985	420	0.53
34	Temperature	temp_1	Araxos	16687001	Lawrimore et al. (2011)	Jan 1951	Dec 1980	360	0.66
35		temp_2	Athens	16714000		Jan 1858	Dec 1975	1416	0.67
36		temp_3	Athens	16714000		Jan 1989	Dec 2001	156	0.68
37		temp_4	Athens	16716000		Jan 1951	Dec 2012	744	0.65
38		temp_5	Heraklion	16754000		Jan 1950	Dec 2015	792	0.69
39		temp_6	Kalamata	16726000		Jan 1956	Dec 2015	720	0.74
40		temp_7	Kerkyra	16641000		Jan 1951	Dec 2016	792	0.67
41		temp_8	Larissa	16648000		Jan 1899	Dec 2016	1416	0.64
42		temp_9	Lemnos	16650000		Jan 1951	Dec 1998	576	0.75
43		temp_10	Methoni	16734000		Jan 1951	Dec 1972	264	0.59
44		temp_11	Methoni	16734000		Jan 1975	Dec 2000	312	0.61
45		temp_12	Patra	16689000		Jan 1951	Dec 1989	468	0.69
46		temp_13	Samos	16723000		Jan 1955	Dec 1969	180	0.64
47		temp_14	Samos	16723000		Jan 1974	Dec 2003	360	0.64
48		temp_15	Souda	16746000		Jan 1961	Dec 2015	660	0.71
49		temp_16	Thessaloniki	16622000		Jan 1892	Dec 2016	1500	0.71
50		temp_17	Thessaloniki	16622001		Jan 1961	Dec 1970	120	0.67

2.4 Methodology outline

We conduct 50 single-case studies by applying a fixed methodology to each time series (see Section 2.1), as explained subsequently. First, we split the time series into a fitting and a test set. The latter is the last observation for the one-step ahead forecasting experiments and the last 12 observations for the multi-step ahead forecasting experiments. Second, we fit the models to the deseasonalized fitting set, within the context determined by each forecasting method (see Section 2.2), and make predictions corresponding to the test set. Third, we add the seasonality to the predicted values and compare them to their corresponding observed using the metrics (see Section 2.3). Finally, we conduct the cross-case synthesis presented in Section 3 to demonstrate similarities and differences between the single-case studies conducted.

3. RESULTS AND DISCUSSION

We visualize the results within and across the individual cases using heatmaps. For the quantitative form of the latter graphs, as well as for Figures S1 and S2, the reader is referred to the Supplementary material, which is available at: http://dx.doi.org/10.17632/p8sw8pzkcd.3.

As regards the heatmaps of the present study, they are formed under the following conditions: a) the darker the colour the better the forecasts and b) the scaling is performed in the row direction. White color rows indicate that no scaling is taking place. The latter happens when the forecasting methods under comparison perform equally well regarding the criterion tested.

In Figures S1 and 1 we present the heatmaps formed for the comparison between the stochastic and two of the ML forecasting methods on precipitation and temperature time series data respectively. As we observe, the results of the single-case studies vary significantly. We also observe that in every individual case examined the following applies. There is no best or worst forecasting method regarding all the criteria set simultaneously. In other words, none of the forecasting methods is uniformly better or worse than the rest. The former observations apply equally to the stochastic and the ML forecasting methods, while it is noteworthy that the Naive benchmark is as competent as the forecasting methods regarding all the criteria set.

The observations outlined above are particularly important, because they reveal that the forecasting quality is subject to limitations. Each forecasting method has some specific theoretical properties and, due to the latter, it performs better or worse than other forecasting methods regarding specific criteria and in specific cases. Thus, the conduct of a single-case study using fewer criteria would have led to a very different overall picture. We note that the metrics RMSE and NSE give almost the same information about the forecast quality regarding the multi-step ahead forecasting experiments, a fact that does not apply to any other pair of metrics.

It is also interesting that the forecasting methods AR(1) and auto_ARFIMA are the least proper to use on precipitation data, while they are competent on the temperature time series data. This is actually a systematic pattern, which can be explained, when tracing back to the single-case studies using precipitation time series data. In more detail, those two forecasting methods predict zero precipitation in contrast to the rest, as a result to the zero precipitation observations in the summer months.

Finally, by studying the numerical results we note that the forecasts for temperature are remarkably better than the forecasts for precipitation. This may be explained by the fact that the variability in temperature is more regular than that in precipitation.

In Figures 2 and S2 we present the heatmaps formed for the sensitivity analysis on the time lags in time series forecasting using the NN and the SVM algorithms respectively. In both figures we observe significant variations in the results across the individual cases, in an extent that it is impossible to decide on a best or worst ML forecasting method among the single-case studies. Regarding the SVM algorithm (Figure S2), we observe no systematic patterns and the variations seem to be rather random, while for the case of the NN algorithm (Figure 2) we observe that the left parts of the heatmaps are smoother with no white cells.

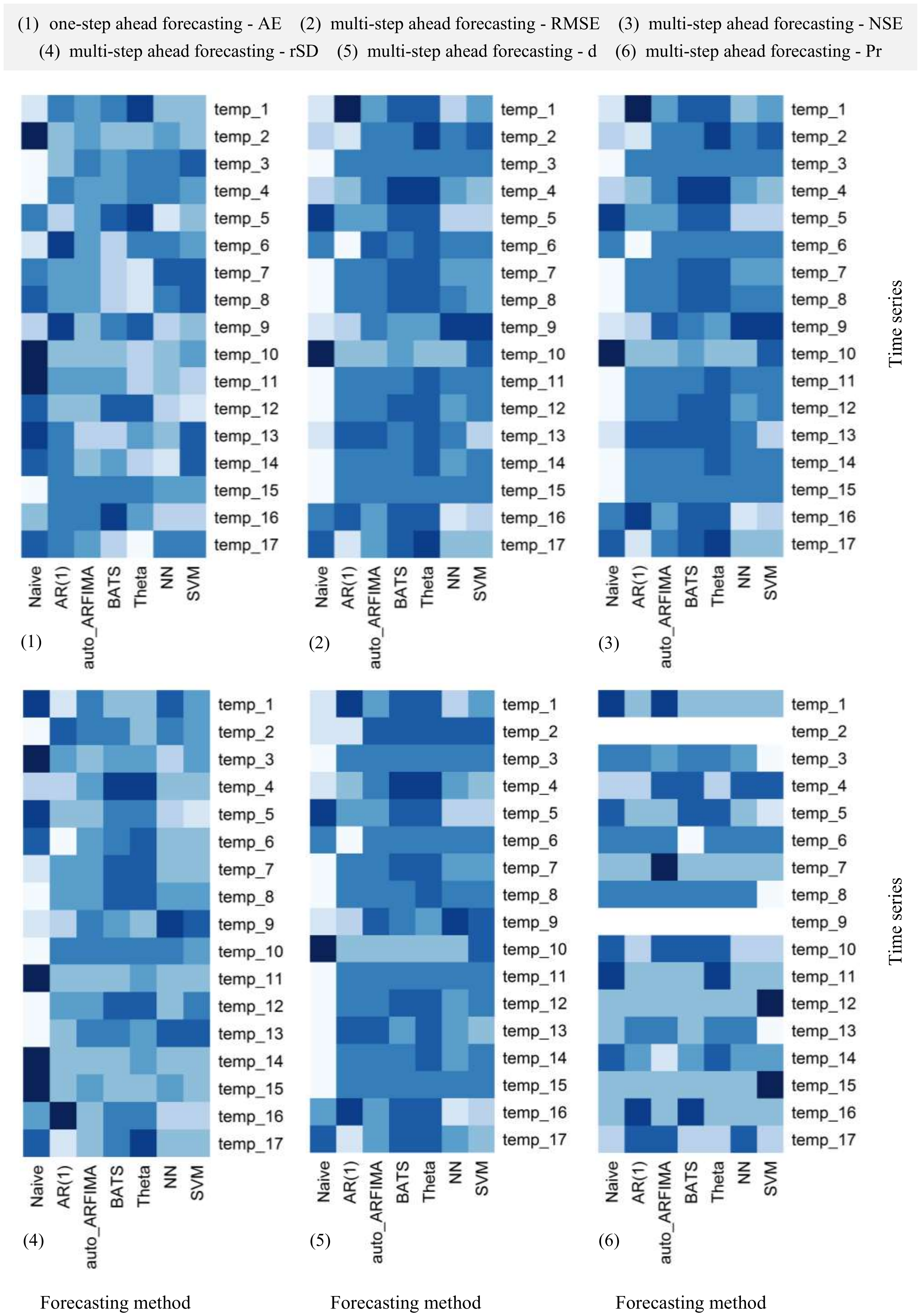

Figure 1. Heatmaps for the comparison between stochastic and ML methods on temperature time series.

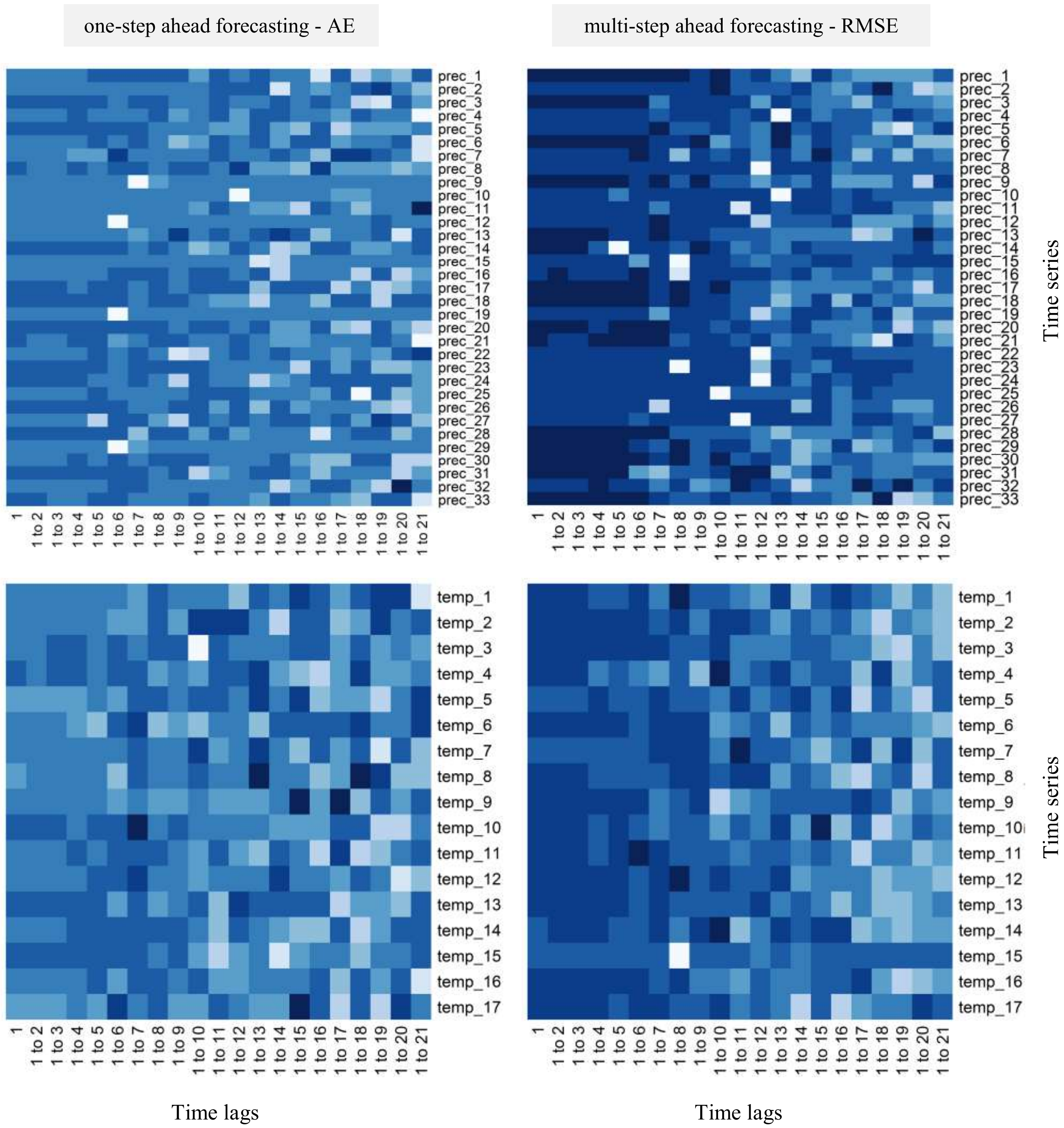

Figure 2. Heatmaps for the sensitivity analysis on the time lags in time series forecasting using the NN algorithm.

In Figure 3 we present the heatmaps formed for the investigation of the effect of the hyperparameter optimization. The results vary across the single-case studies in a rather random manner, which indicates that the hyperparameter optimization does not necessary lead to better forecasts for the NN and SVM algorithms.

4. CONCLUSIONS

The multiple-case study conducted must be encountered as a contingent empirical evidence on several issues that have drawn the attention in the field of time series forecasting. The findings suggest that the stochastic and ML methods can perform equally well, but always under limitations. The best forecasting method depends on the case examined and the criterion of interest, while it can be either stochastic or ML. However, the ML methods are computationally intensive. Regarding the time lag selection, the best choice seems to depend mainly on the case, while the ML algorithm might has also some effect. Finally, for the algorithms used in the present study hyperparameter optimization does not necessarily lead to better forecasts.

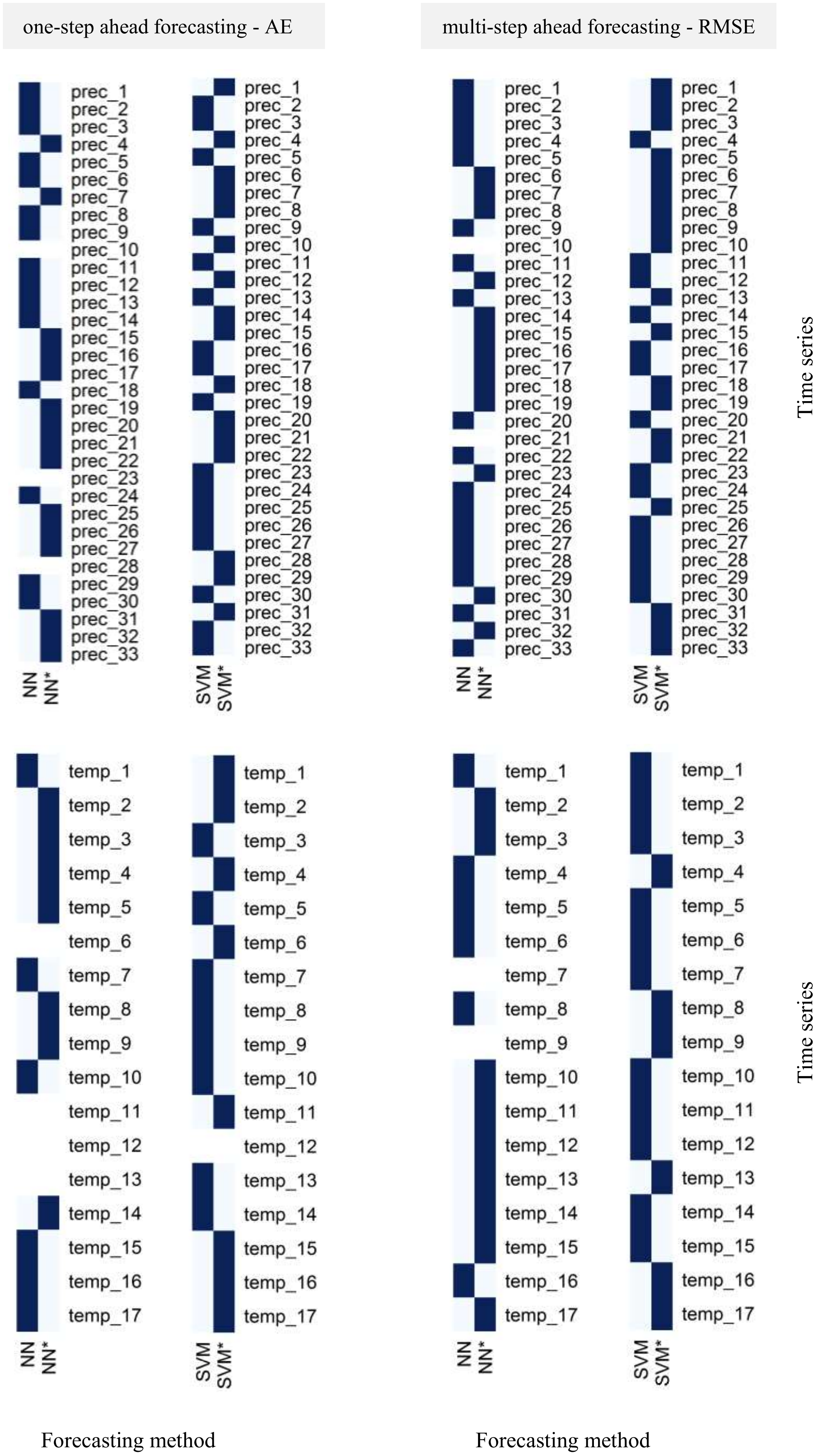

*Figure 3. Heatmaps for the investigation of the effect of hyperparameter optimization on the forecast quality. The symbol * in the name of a forecasting method denotes that the model's hyperparameters have been optimized.*

REFERENCES

Achen, C. H., Snidal D., 1989. Rational deterrence theory and comparative case studies. World Polit.; 41(2): 143-169. doi:10.2307/2010405

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn.; 20(3): 273-297. doi:10.1007/BF00994018

Cortez, P., 2010. Data mining with neural networks and support vector machines using the R/rminer tool. In: Perner P (eds) Advances in Data Mining. Applications and Theoretical Aspects. Springer Berlin Heidelberg, pp 572-583. doi:10.1007/978-3-642-14400-4_44

Cortez, P., 2015. rminer: Data mining classification and regression methods. R package version 1.4.1. https://cran.r-project.org/web/packages/rminer/index.html

Hyndman, R. J., O'Hara-Wild, M., Bergmeir, C., Razbash, S., Wang, E., 2016. forecast: Forecasting functions for time series and linear models. R package version 7.1. https://CRAN.R-project.org/package=forecast

Hyndman, R. J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. J. Stat. Softw.; 27(3): 1-22. doi:10.18637/jss.v027.i03

Koutsoyiannis, D., Yao, H., Georgakakos, A., 2008, Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. Hydrolog. Sci. J.; 53(1). doi:10.1623/hysj.53.1.142

Krause, P., Boyle, D. P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci.; 5: 89-97

Lawrimore, J. H., Menne, M. J., Gleason, B. E., Williams C. N., Wuertz, D. B., Vose, R. S., Rennie, J., 2011. An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. J. Geophys. Res.; 116(D19121). doi:10.1029/2011JD016187

Maier, H. R., Dandy, G. C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Model. Softw.; 15(1): 101-124. doi:10.1016/S1364-8152(99)00007-9

Papacharalampous, G. A., Tyralis, H., Koutsoyiannis, D., 2017. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. Journal of Hydrology. In review

Peterson, T. C., Vose, R. S., 1997. An overview of the Global Historical Climatology Network temperature database. Bull. Am. Meteorol. Soc.; 78(12): 2837-2849. doi:10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2

Raghavendra, N. S., Deka, P. C., 2014. Support vector machine applications in the field of hydrology: a review. Appl. Soft Comput.; 19: 372-386. doi:10.1016/j.asoc.2014.02.002

Taieb, S. B., Bontempi, G., Atiya, A. F., Sorjamaa, A., 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert Syst. Appl; 39(8): 7067-7083. doi:10.1016/j.eswa.2012.01.039

Tyralis, H., 2016. HKprocess: Hurst-Kolmogorov Process. R package version 0.0-2. https://cran.r-project.org/web/packages/HKprocess/index.html

Tyralis, H., Koutsoyiannis, D., 2011. Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. Stoch. Env. Res. Risk.; 25(1): 21-33. doi:10.1007/s00477-010-0408-x

Valipour, M., Banihabib, M. E., Behbahani, S. M. R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J. Hydrol.; 476: 433-441. doi:10.1016/j.jhydrol.2012.11.017

Vapnik, V. N., 1995. The nature of statistical learning theory, first edition. Springer-Verlag New York. doi:10.1007/978-1-4757-3264-1

Vapnik, V. N., 1999. An overview of statistical learning theory. IEEE Trans Neural Netw; 10(5): 988-999. doi:10.1109/72.788640

Zambrano-Bigiarini, M., 2014. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-8. https://cran.r-project.org/web/packages/hydroGOF/index.html

Zeileis, A., Grothendieck, G., 2005. zoo: S3 infrastructure for regular and irregular time series. J. Stat. Softw.; 14(6): 1-27. doi:10.18637/jss.v014.i06